DOCUMENT RESUME

ED 267 096 TM 860 172

AUTHOR Greeno, James G.; And Others

TITLE Investigations of a Cognitive Skill.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.; Office of Naval Research, Arlington, Va.
REPORT NO LRDC-1985/27
PUB DATE 85
CONTRACT NR-667-430
NOTE 40p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Algebra; *Cognitive Processes; Elementary Secondary Education; *Error Patterns; Higher Education; Language Processing; *Learning Theories; Mathematical Logic; Mathematical Models; *Mathematics Skills; *Models; *Pattern Recognition; Problem Solving; Secondary Education

ABSTRACT

Research was conducted investigating properties of skill in learning, in the domain of elementary algebra. Thinking-aloud protocols indicate that early knowledge of the subjects studied was fragmentary, rather than involving systematically flawed procedures. Computational models, developed to simulate observed errors, focused on the role of structural representations in facilitating reliable performance. Connectionist models for recognizing structural features were investigated, leading to the conclusion that the cognitive system probably requires knowledge functionally equivalent to grammatical rules. Data from information processing experiments indicated that: (1) judgments about the application of an algebraic operator are influenced by low-level features recognized before a completely parsed representation is formed; and (2) recognition of individual characters in expressions is not facilitated by syntactically correct contexts, as it is by lexical contexts in letter recognition, but information about the algebraic categories of characters is obtained early in processing from the syntactic context. The conclusion is made that training in basic symbolic skill might be more effective if more attention were given to teaching the structure of information of the domain, including general features of the information presented in problems as well as general constraints and goals of the procedures to be acquired. (Author/LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267096

LEARNING RESEARCH AND DEVELOPMENT CENTER

INVESTIGATIONS OF A COGNITIVE SKILL

1985/27

JAMES G. GREENO, MARIA E. MAGONE, MITCHELL RABINOWITZ, MICHAEL RANNEY, CLAUSS STRAUCH, AND THERESA M. VITOLO

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Aug

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 172

University of Pittsburgh

INVESTIGATIONS OF A COGNITIVE SKILL

James G. Greeno, Maria E. Magone, Mitchell Rabinowitz,
Michael Ranney, Clauss Strauch, and Theresa M. Vitolo

Learning Research and Development Center
University of Pittsburgh

1985

The research reported herein is part of the final report of a project entitled, "Principles of Cognitive Skill and Acquisition," for which James G. Greeno and Lauren B. Resnick were co-principal investigators. The project was included in Contract NR 667-430, Office of Naval Research. The research was conducted at the Learning Research and Development Center, funded in part by the National Institute of Education, U.S. Department of Education. The opinions expressed do not necessarily reflect the position or policy of the above agencies, and no official endorsement should be inferred. Many of the results reported were presented in a paper given by Greeno at the meeting of the American Psychological Association, Anaheim, CA in August of 1983.

Investigations of a Cognitive Skill

James G. Greeno, Maria E. Magone, Mitchell Rabinowitz,
Michael Ranney, Clauss Strauch, and Theresa M. Vitolo

ABSTRACT

Research was conducted investigating properties of skill in learning, in the domain of elementary algebra. Thinking-aloud protocols indicate that early knowledge of the subjects studied was fragmentary, rather than involving systematically flawed procedures. Computational models, developed to simulate observed errors, focused on the role of structural representations in facilitating reliable performance. Connectionist models for recognizing structural features were investigated, leading to the conclusion that the cognitive system probably requires knowledge functionally equivalent to grammatical rules. Data from information-processing experiments indicated that (a) judgments about the application of an algebraic operator are influenced by low-level features recognized before a completely parsed representation is formed; and (b) recognition of individual characters in expressions is not facilitated by syntactically correct contexts, as it is by lexical contexts in letter recognition, but information about the algebraic categories of characters is obtained early in processing from the syntactic context. The authors conclude that training in basic symbolic skill might be more effective if more attention were given to teaching the structure of information of the domain, including general features of the information presented in problems as well as general constraints and goals of the procedures to be acquired.

The research reported here investigated properties of a cognitive skill in the early stages of its acquisition. The studies focused primarily on performance of students who were taking their first course in elementary algebra.

Properties of early skill were investigated using several methods. First, general characteristics of performance were studied by obtaining thinking-aloud protocols from students working on algebra problems. Eight students in ninth-grade beginning algebra courses volunteered to be interviewed approximately once per week during the first semester of their study of algebra. Each interview lasted about 20 minutes. In most interviews, students solved a few problems of the kind they had in homework during that part of the course. Additional questions and some unusual problems were also included to assess students' understanding of some general concepts. Protocols were recorded on audio tape and transcriptions were made with the students' paper-and-pencil work coordinated with the verbal data.

A second research activity was the construction of computational models to simulate some significant aspects of the students' performance. Based on the protocol data, we concluded that the early form of skill in this domain is best characterized as a loosely organized set of fragments, rather than a systematic structure of procedural knowledge. Our modeling effort investigated questions about how fragmentary knowledge produces performance in a symbolic domain. We focused on the issues of comprehension, asking what the cognitive requirements are for a system to achieve structural representations of grammatical expressions.

A third research activity was conducting experiments in which we studied perception of characters in algebra expressions exposed for brief periods and measured latencies of judgments about algebraic expressions based on structural properties of the expressions. Results of these studies provide information about some characteristics of basic information processing involved in the cognitive skill of algebra.

Finally, in research related to the studies reported here, we have investigated relations of new skill acquired in the study of algebra to the students' previous knowledge. This has included interviews with students before they began their study of algebra, investigating their understanding of relations between arithmetic notation and quantitative operations (Chaiklin & Lesgold, 1984). We also have developed some new instructional tasks to provide background knowledge relevant to learning algebra that we have concluded was relatively weak or absent in the students whose performance we observed. These studies of prerequisite knowledge will not be discussed in this report, but their results are consistent with the general conclusion that skill early in learning algebra is fragmentary and unsystematic.

1. Fragmentary Nature of Early Skill

The background for this research is provided by recent analyses of cognitive skill, especially in the domain of mathematics. Analyses in domains apparently similar to algebra have been provided. Problem solving in geometry has been analyzed by Anderson (1982) and by Greeno (1978). Performance in arithmetic has been analyzed by Brown and Burton (1978) and by Groen and Resnick (1977). In both of these cases, the skills that students acquire appear to be quite systematic. Models that simulate

students' performance include significant control structure and strategic knowledge that organizes problem-solving activity.

With this background, the main findings of our research were unexpected. Instead of observing performance that was systematic, apparently governed by a coherent control structure, we found performance that was profoundly disorganized and fragmentary.

The evidence for this conclusion is primarily in the nature of errors that we observed in the protocols we obtained from beginning students. The errors were unsystematic, of the kind that have been called "slips," (VanLehn, 1981) rather than being caused by "bugs," or procedural flaws that cause performance that is wrong in systematic ways.

Figure 1 shows an example of a student's writing on a large but otherwise simple problem. Things went well until Line 4, where 15a + 16 was transformed into 31a. Then in Line 5, 5[31a] was transformed into 155. These errors could be produced by systematic flaws in a student's procedural knowledge, but apparently they were not. For example, the error of combining two terms like 15a and 16, where only one of them includes a variable, could have occurred in the transformation from Line 1 to Line 2 -- that is, 6a + 8 could be transformed to 14a. Indeed, another student working on this same problem did perform that transformation. But the error did not occur systematically, and most of the errors we observed were unsystematic in this way.

Simplify: $2a + 5\,[(6a + 8)2 + 3a]$

$2a + 5\,[12a + 16 + 3a]$

$2a + 5\,[15a + 16]$

$2a + 5\,[31a]$

$2a + 155$

Figure 1. Examples of errors apparently caused by "slips," typical in early stages of skill acquisition.

We analyzed the errors in our complete set of data to examine their systematicity. We assigned to each error a characterization that could constitute a bug, in the sense of Brown and Burton (1978). For example, the error in Line 4 of Figure 1 could result from a procedure in which a variable in one term is noticed, and that term is combined with another term unless the second term has a different variable.*1

Table 1 shows a summary of this analysis. For each error type that we characterized for a given student, we examined that student's performance on the problems in the same interview session as the error or errors of that type. We counted the occasions in that session on which the error would have occurred according to our characterization if the error had resulted from a systematically flawed procedure. The number of those occasions, including the error(s), is called the number of opportunities. We applied a threshold of errors occurring on 0.5 of the opportunities, and

---

*1. There is an unavoidable arbitrariness in this analysis, because the appropriate characterizations of errors cannot be identified uniquely. The characterization of the Line-4 error could be more specific -- e.g., applying only when a term with a variable comes first and a term without a variable follows it immediately -- or it could be more general -- e.g., applying to all pairs of terms, whether they have variables or not. The characterizations we used reflect our judgments of the plausibility of systematic flaws that could have caused the errors.

restricted attention to error types that had at least three opportunities. As Table 1 shows, the majority of the error types that we observed did not occur on more than 0.5 of their opportunities. Most of the error types that were systematic involved exponents. (For example, one student systematically simplified terms like $3x^2$ by multiplying the coefficient and the exponent to get 6x.) If expressions with exponents are excluded, then 18 of the 22 error types in our data occurred on 0.5 or fewer of their opportunities.

Table 1

Numbers of Error Types with Errors Given by Individual Students on $>.5$ and $\leq .5$ of Opportunities.

	Error Types with $>.5$ Errors	Error Types with $\leq .5$ Errors
Expressions without Exponents	4	18
Expressions with Exponents	17	11
Total	21	29

Note: Only error types with ≥ 3 opportunities and ≥ 1 errors are included.

There were a few quite dramatic examples of performance that seem to result from fragmentary procedural knowledge. We present part of one such example. The task was to solve the following equation:

$$3y + 8 = 2y - y$$

The student said, I guess you take the positive 8, make it negative 8, and

make the y, positive y. He wrote "-8" and "+y" on the paper, as follows:

$$3y + 8 = 2y - y$$
$$\quad -8 \qquad\quad +y$$

These steps could be included in legal operations, but the student seemed to lack strategic knowledge of what they are for or knowledge of constraints on their use. Next, the student said, "Bring this negative 8 down here, too, I guess," writing another "-8" on the second line under "3y." We suppose that this might have been caused by knowledge that quantities should be subtracted twice in solving an equation, although the requirement of subtracting on the two sides of the equal sign was not observed. The student continued on in a persistent albeit quite haphazard way, until the following display had been created:

$$3y + 8 = 2y - y$$
$$-8 \quad -8 \qquad +y$$
$$\text{---------------}$$
$$-5y \quad 3y \quad 2 \quad 2$$
$$-2y \quad = 1y \quad = 1y$$
$$-2y \quad = \quad -2y$$

Performance like this contrasts sharply with performance of students early in their study of geometry, which Greeno (1978) observed in a study similar to this one. Not all the geometry students could solve all the problems correctly, of course. However, when they had not acquired the knowledge they needed, they usually did not do anything, saying "I don't know" or "I'm stuck," rather than proceeding to perform inappropriate operations as was typical of the algebra students. We consider it possible that acquisition of problem-solving skill in geometry occurs quite differently from acquisition of algebra, with earlier learning of an appropriate control structure. Geometry differs from algebra in several

ways that might make that happen; for example, proof exercises present specific goals to be achieved, and the problems include diagrams that present patterns that are closely related to the operations to be performed. Of course, other differences could have caused the difference between our studies, including differences between the students whom we observed.

Performance of our algebra students also contrasts sharply with the hypothesis incorporated in the BUGGY system (Brown & Burton, 1978), that errors are caused by variants of a coherent procedural network. More recent analyses of subtraction errors (Brown & VanLehn, 1980; VanLehn, 1983) have used a view more consistent with our findings, that errors occur because of incomplete knowledge, and that students "repair" their procedures with local problem-solving heuristics when they encounter situations for which their knowledge is inadequate. However, the degree of incompleteness that characterizes the students we observed is so extreme that it seems more accurate to model their knowledge as a collection of disconnected fragments than as a structure that is well organized but incomplete.

Our findings are at variance from those obtained for algebra in other studies by Carry, Lewis, and Bernard (1980), by Davis, McKnight, and Jockusch (1978), by Matz (1982), and by Sleeman and Smith (1981). These investigators have focused on more systematic aspects of performance, such as misinterpretations of verbal descriptions of procedures and operators that are consistently applied in an overgeneralized way. Our students' performance probably was less systematic than that observed by Carry et al and by Sleeman and Smith, partly because we observed students as they were acquiring the procedures initially, and we gave only a few problems in each

interview session. In addition, the emphasis that we place on unsystematic errors is complementary to the emphasis given in other analyses to systematic errors, since both undoubtedly occur and need to be understood theoretically. Indeed, our observations include a few quite systematic errors, most involving expressions with exponents, that probably can be understood as instances of repairs and resulting mal-rules.

2. Models of Errors

We have conducted a theoretical investigation of some of the kinds of errors that occurred in our data. This modeling investigated the role of structural information in the occurrence of errors.

Students have to learn to parse expressions -- that is, to recognize structural features such as terms made up of coefficients, variables, and exponents, and subexpressions composed of sets of terms and operators. Correct use of algebraic operators depends on structural features; for example, the operation of combining terms can be applied to simplify the expression 3X(5Y-2Y)+7Z but not to 3X(5Y+7Z)-2Y. However, students' skill in parsing expressions may be only partially developed when they begin to acquire knowledge of the operations.

To investigate the possible role of parsing knowledge, we formulated two models of correct performance of some operations, and degraded each of them to simulate errors that we observed in students' performance. In one version of correct performance we assumed that representations of expressions include correct structural features, such as coefficients, variables, terms, and subexpressions. The representation of an expression is a tree with each node representing a subexpression, term, operator, coefficient, or variable. Knowledge of the operations was represented as

sets of productions, and the conditions of the productions referred to the structural properties. For example, a condition for combining terms is that two terms with the same variable or sequence of variables must be included in a subexpression at a single level of the representation. We call the models with this representation the models with structure.

In the other version of correct performance the representation of expressions was nonstructural and linear. Characters in the expression are distinguished by category as numerals, letters, operators, and parentheses, and the only structural feature is left-to-right linear order. We call these models without structure.

The general finding of this modeling effort was that models with structure require much more substantial change to degrade them so they simulate errors than do models without structure. Degrading models without structure involved removing a production or removing a feature from a tested condition, changes that seem quite plausible as causes of "slips." Degrading the models with structure required changes such as redefining the conditions for performing an action, replacing a set of features with structurally weaker features.

An example is the error of inappropriately combining terms, such as Line 4 of Figure 1. In the correct model with structure, the procedure is defined on the structural components, and a check is included that the variables of terms are matched. To degrade the model to make the error, five alterations are needed, including a change from testing the variables to testing the terms, dropping a subtest for differences, and changing the operation so that it applies to numbers rather than coefficients.

In the correct model without structure, the procedure is defined on number-and-letter sequences. Degrading this model requires only removing some productions that test letters, rather than changing their features. The operations that are performed do not need to be changed.

A second kind of error that we simulated involves signs of terms. For example, "-15x + -24x" was transformed to "39x" by one student, and "2y - y" was transformed to "3y" by another student. In the correct model with structure, the signs of terms are represented as components of the terms, and the model is degraded by removing components that are integral parts of the processing of terms. In the correct model without structure, signs of terms are just symbols that happen to precede numerals or letters in the spatial array, and degradation involves removing components of the procedure for processing the symbols for signs that are unrelated to other components of the procedure.

A third error type that we considered involves dropping a variable when multiplication is performed. For example, "-8(4 - 3d)" was transformed to "-32 + 24" by one student. This kind of error can be simulated by removing detection of the letter from either the correct model with structure or the correct model without structure. Another alternative, though, is that in the model with structure, a subgoal to process the variable of a term is omitted, and in the model without structure the processing of the letter is omitted. In this version, the change needed with structural features removes a component that is included in an integrated procedure, while the change without structure is a separate action.

Our conclusion is that the structural features provided by comprehension processes may make performance more reliable because it provides information units that are needed for cognitive procedures to be integrated and organized. Conversely, errors of the kind that we observed probably indicate that students' processes of representing expressions do not provide them with well-formed representations of the structural features of the expressions during their early stages of learning.

3. Models of Parsing

The generally unsystematic character of performance that we observed raises an interesting problem of modeling the skill. A major characteristic of most information-processing models is that performance depends strongly on the model's control structure. Detailed analyses have been provided about both general problem-solving strategies such as means-ends analysis (Newell & Simon, 1972) and domain-specific knowledge for planning (Sacerdoti, 1977).

The performance that we observed suggests that the knowledge of students in early learning of algebra lacks a coherent control structure. The question that arises, then, is how to construct a model that simulates their performance. Normally, if we write a computer program in which the control structure is faulty, the program will not run at all. Students in algebra, however, almost always do something -- their "programs" continue to run, albeit incorrectly, rather than halting.

There seem to be two general theoretical alternatives available to us. Using production systems, erratic performance can be simulated with partial matching or multiple productions, so that different productions will be executed on different occasions. John Anderson's (1983) ACT model simulates variability in this way, and our analyses of errors described in Section 2 uses this approach.

A somewhat more radical approach to modeling variable performance is also available, and we used it to simulate processes of parsing expressions. This approach uses a framework called connectionism, being developed by investigators such as James Anderson (Anderson, Silverstein, Rity, & Jones, 1977), Feldman and Ballard (1982), Hinton (1981), and McClelland and Rumelhart (1981).

An Issue Raised by Connectionism

Use of the connectionist framework to model parsing enables us to address a fundamental issue in cognitive theory, introduced by Chomsky (e.g., 1965). The question is whether generative symbolic behavior can be achieved by a system that lacks primitive symbolic processes. Chomsky argued that to account for understanding and production of novel sentences, it is necessary to assume that individuals have implicit knowledge of grammatical rules, which he called competence. Chomsky's arguments were directed specifically against behaviorist and associationist theories in which knowledge is limited to undifferentiated connections between stimuli and responses, or between ideas. Newell and Simon (e.g., 1976) also have articulated the view that symbolic operations are primitive cognitive processes.

The issue, as we understand it, is as follows. Generative performance is observed in symbolic domains, including language but also generally in problem solving. By "generative performance" we mean performance that cannot be explained on the basis of specific actions that are associated with specific stimulus conditions. Instead, individuals perform in ways that are consistent with general rules that are formulated on classes of situations and actions. The specific performances that individuals display are extremely variable, and include instances that are completely novel at the level of specific situations and actions, so it is not possible to account for their performance by assuming that they are based on specific situation-action associations.

The position taken by Chomsky, Newell and Simon, and others is that we must attribute knowledge of general rules to individuals whose performance is generative. We believe that the critical property of this knowledge is that it involves transmission of symbolic information between components of the cognitive process. Mental states are characterized according to symbolic information that they include -- for example, a word may or may not have been recognized, or a noun phrase may or may not have been represented. It is assumed that the specific information included in one state causes the information that is included in other states. For example, recognition of a sequence of words causes representation of a phrase. The processes by which information states are causally connected correspond to the rules that the individual knows, albeit the knowledge may be implicit (and very often is).

A critical feature of the rules, enabling generative performance, is that they involve general classes of symbolic structures rather than specific word sequences. For example, a noun phrase will be represented when the sequence of words is Det Adj Cnoun where Det is any determiner, Adj is any adjective, and Cnoun is any common noun. A noun phrase is represented for "the furious brick," even though that specific sequence of words has never been encountered.

The claim of connectionism (formerly behaviorism or associationism) is that generative performance is an emergent property, resulting from non-symbolic cognitive mechanisms. In a connectionist theory, the cognitive system consists of a fixed set of units, each of which varies in its level of activation. Units are connected to other units, and the connections transmit excitation and inhibition between pairs of units. The important constraint is that transmission of symbolic information is not permitted (beyond the activation levels, which may be thought of as "information" if one likes, but are not symbolic in the usual sense). A state of the system is just the collection of activity levels of all its units.

Discussions of the adequacy of connectionist models in the 1960s (e.g., Dixon & Horton, 1968) led many psychologists to the view that symbolic processes are required to account for complex phenomena of language and problem solving. Recently, however, connectionism has reappeared in a more complex and sophisticated form than it had 20 years ago. The issue can be addressed again, and perhaps the outcome will be different.

Connectionist models have been focused primarily on phenomena in pattern recognition (e.g., Anderson et al, 1977), word recognition (McClelland & Rumelhart, 1981), and other phenomena in which specific patterns of features are recognized. The simplest hypothesis for these phenomena is that there are units in the cognitive system that correspond to the patterns that can be recognized. For example, McClelland and Rumelhart's model includes a distinct unit for each word in the vocabulary. Successful recognition of a word occurs when that word's unit is sufficiently active, where "sufficiently" means exceeding a threshold. The activation of that unit is increased by the activity of other units that recognize the letters of the word.

In parsing a sentence or an algebraic expression, patterns are recognized that do not correspond to known patterns -- for example, recognition of "the furious brick" as a noun phrase, or "$8xy^2$" as a term, requires use of general structural properties rather than specific sequences of characters. Our theoretical effort, then, was to try to understand the kinds of connectionist structures that could produce representations of syntactic structure in a generative way, and to use the results to re-evaluate the question of whether symbolic processes are fundamental components of cognitive systems.

Properties of Models*2

We have been able to find two kinds of connectionist models that construct representations of structure. We have written programs that implement a few versions of one kind of model.

---

*2. We are grateful to Geoffrey Hinton and James McClelland for discussions about the models that we discuss in this section.

The models that we have implemented include a process that generates new nodes and connects them in the cognitive network. These models include virtually no structural knowledge of algebra, but generate nodes on the basis of weak spatial features and select nodes using connections that are differentiated only on the basis of categories of individual characters.

The other kind of model has no process of generating new nodes. Its cognitive units are organized into modules that are specialized for recognition of types of patterns. These pattern modules are linked by mapping units that cause patterns of activation in one module to produce patterns in another module.

Regarding the general issue of symbolic processes, the second type of model -- the one with pattern modules -- has structures that we interpret as direct implementations of syntactic rules. For example, a standard grammar for parsing algebraic expressions would include the rule: "Term" --> "Numeral" + "Variable" in some form. A symbolic parser using this rule would recognize a unit consisting of a numeral followed by a variable (such as "3x") as a term. In the connectionist models that we have been able to conceptualize, there are specialized modules that recognize the characters (e.g., "3" and "x"), and another module, activated by the character-recognizing modules, that becomes active because the characters recognized are the correct sequence of types (e.g., a numeral followed by a variable). Therefore, although these models do not include the rules of a grammar explicitly, the modules that they contain, and the activation sequences that occur, implement knowledge of rewrite rules in a fairly straightforward manner.

The models that we implemented do not have pattern modules, and they do not have knowledge that corresponds directly to grammatical rules. However, their ability to generate new structural components contradicts the connectionist constraint of having a fixed set of structural components that vary only in the parameters of activation. Our conclusion is that if a model for parsing is restricted so that it cannot produce new structure, it probably has to include mapping structures that are the functional equivalent of grammatical rules for rewriting symbolic information structures.

Models that generate new units. One requirement of a parser is that it can recognize constituent units that have the correct structure but are not specifically known. For example, a parser for algebra should recognize $8xy^2$ as a term with a structure like that shown in Figure 2. The constituent units are the coefficient 8, separated from the variable part of the expression, which is divided between x and y^2. The parser must be able to produce a representation like Figure 2 without prior knowledge of the specific units that it will encounter.

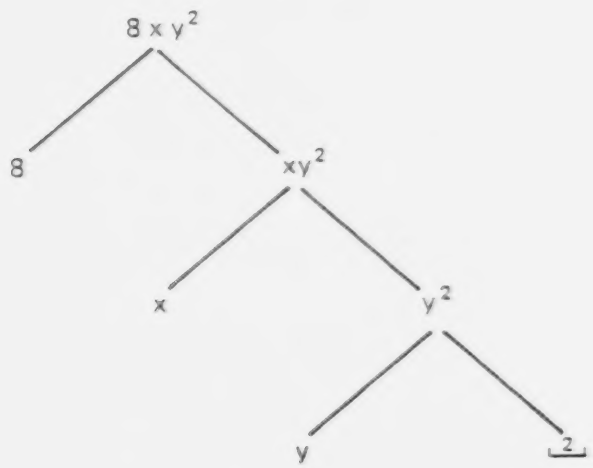

Figure 2. A structural description of a term.

An obvious way to achieve this is to have a process that generates nodes corresponding to the constituent units of the expression. This is what ordinary parsing systems do, with the nodes generated according to grammatical rules (e.g., Det+Adj+Cnoun-->Nphrase). The systems that we constructed do not generate nodes according to grammatical rules, but rather on the basis of weak spatial properties. Nodes are generated corresponding to combinations of characters that are in an appropriate sequence. A weak version forms nodes from pairs that simply are in the correct left-to-right sequence, even if other characters intervene between them, so that for $8xy^2$ there would be nodes generated for 8x, 8y, 8^2, xy, x^2, and y^2. Based on pair-nodes that achieve a threshold of activity, nodes for triples are generated, such as (8x)y, $(8x)^2$, $(8y)^2$, 8(xy), $8(x^2)$, $8(y^2)$, and $x(y^2)$. Nodes for pairs of pairs are also generated, such as $(8x)(y^2)$. Then quadruples are represented, such as $((8x)y)^2$, $(8(xy))^2$, and $8(x(y^2))$. A slightly stronger version only generates nodes for sets of characters that are adjacent in the expression.

Nodes that are included in the network are connected to the nodes for their constituents, and excitation from the lower-level nodes increases activation of the higher-level nodes. Higher-level nodes may be connected to each other and transmit inhibition, thus producing a kind of competition.

Figure 3 shows the network produced in processing the expression "3xy." Excitatory links are indicated with arrow heads and inhibitory links are indicated with dots. In this version, each higher-level node inhibits other higher-level nodes that are less complex than it is, but it does not inhibit its own constituents.

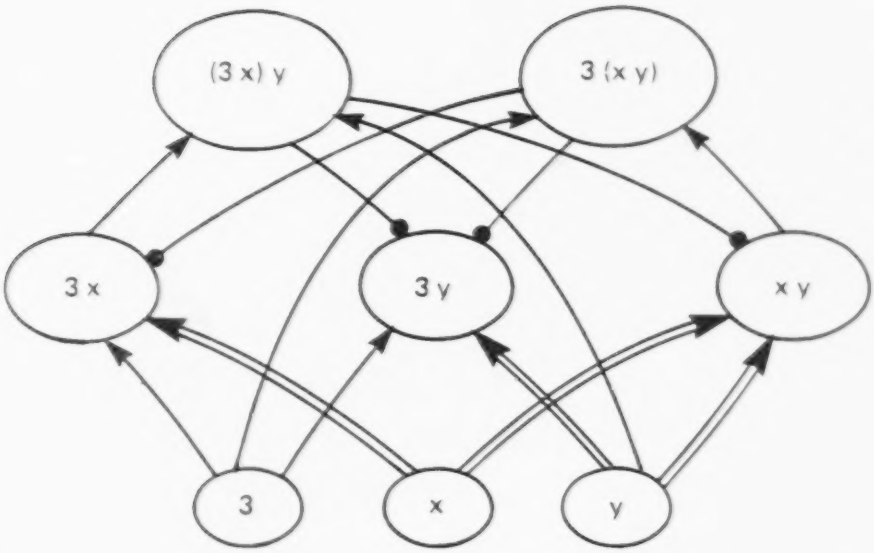

Figure 3. Structure generated with nodes for constituent units and selection based on different strengths from characters based on categories.

We considered the question of selection of a "correct" structural representation. Students should learn, for example, that the two main constituents of "3xy" are the coefficient "3" and the variable sequence "xy." Thus, the structure 3(xy) is preferred to (3x)y in Figure 3. We found quite a simple way to arrange the model to select a preferred structural description. This involved variations in the strengths of links in the network.

We allowed the strength of excitatory links from single characters to higher-level components to vary according to the categories of the characters. The variations we used involved giving letters greater strength than ordinary numerals, and superscript numerals (i.e., exponents) greater strength than letters. In Figure 3 this is indicated by the double arrows from "x" and "y" to the second-level nodes that contain them.

An alternative way to achieve selection of a preferred structure involves a kind of generic lexicon. It might be that students acquire cognitive units corresponding to sequences of characters, represented at the level of their categories. Such units would correspond to schemata that would be instantiated when sequences of the correct kind are encountered. Examples of such sequences would be <Num+Let>, for a numeral followed by a letter such as "3x," or <Let+Let> for a pair of letters such as "xy," or <Num+(Let+Let)> for a numeral and a pair of letters such as "3xy." Figure 4 shows a network for "3xy" that includes generic lexical units. Selection of the preferred structure 3(xy) rather than (3x)y is assured if the <Let+Let> unit transmits excitation to its instances more strongly than <Num+Let> does.

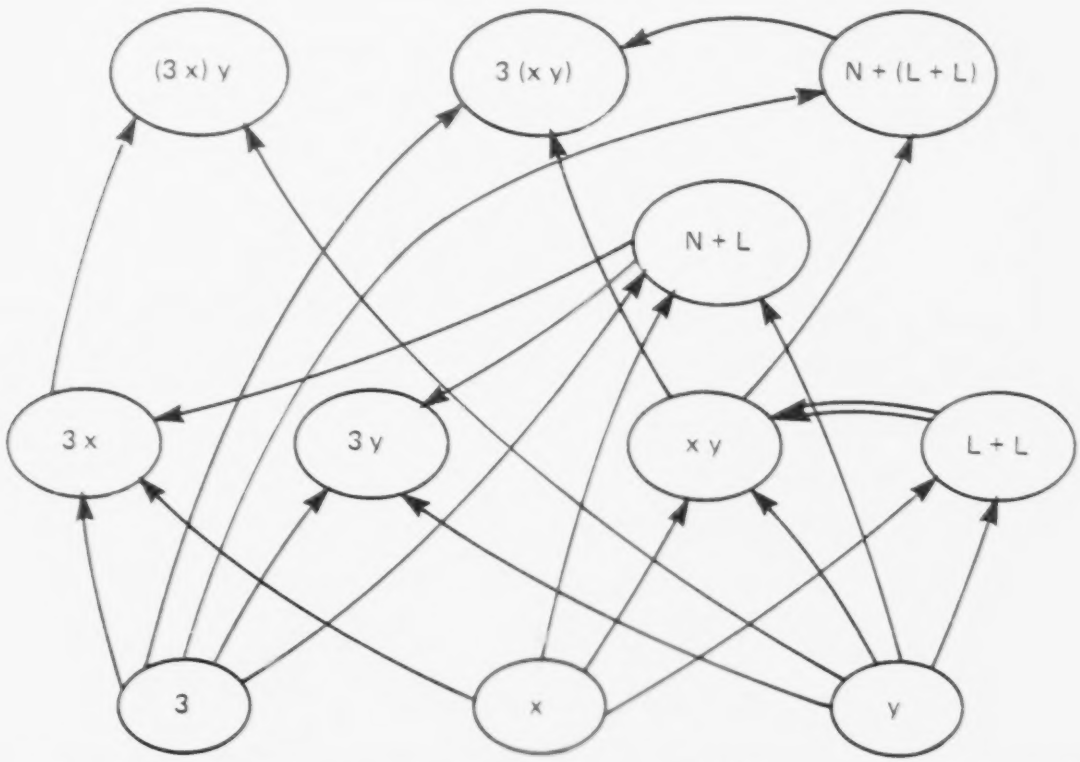

Figure 4. Structure with selection based on different strengths of facilitation from nodes in a "generic lexicon."

Pattern-recognizing modules. A completely different hypothesis is that patterns are recognized by modules of cognitive units, rather than by individual units. The recognition of a pattern by a module corresponds to a configuration of activation of its elements. Some of the elements in the module carry categorical information, so patterns can be recognized on the basis of structural features. Hinton (1981) has implemented an illustrative system of this kind, which is capable of recognizing patterns with the structure "Agent+Action+Object."

In addition, this kind of system requires mapping units, which can cause distinctive patterns in one module based on patterns in other modules. When the mapping units are sensitive to category-based elements the result is a system in which structural descriptions can be generated. A sketch is shown in Figure 5, where [x] and [y] refer to patterns that include information that these are letters, and [xy] is a pattern that is represented because [x] and [y] are represented, through a set of mapping units. Then a pattern corresponding to [3(xy)] is represented through the joint activation of [3] and [xy] because of another set of mapping units.

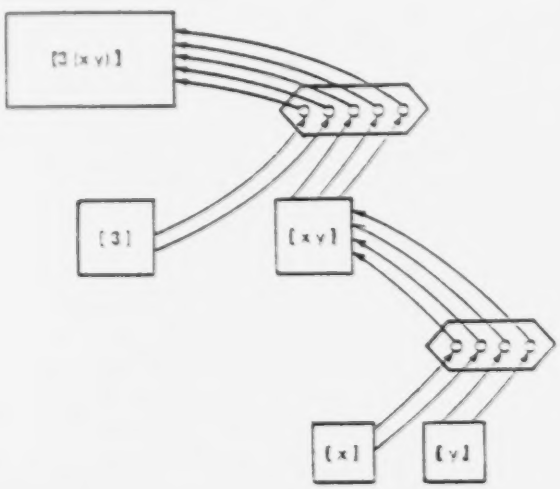

Figure 5. Patterns recognized with structure-specific mapping units.

Implementation of this kind of system is beyond the capabilities of the computational resources that we have had available for the project, but it seems feasible. It has the advantage that it does not require generation of structure in order to form structural representations. On the other hand, the mapping units that it includes are functionally equivalent to the rewrite rules of a grammar. For example, the units that form a pattern "Letter + Letter" from two patterns "Letter" and "Letter" are equivalent to the recognizer for a rule: "Variable String" --> "Letter" + "Letter." The other set of units in Figure 5 is equivalent to a recognizer for the rule: "Term" --> "Numeral" + "Variable String." This seems to confirm the claim that generative performance requires symbolic processing as a primitive component of a cognitive system.

Use of spatial information. A source of information that is potentially useful for parsing expressions is their spatial layout. Characters that form a term are spatially contiguous and are separated from characters in other terms by operators. Subexpressions are located with punctuation marks such as parentheses and fraction bars.

It is reasonable to hypothesize that spatial information plays a role in the comprehension of algebra expressions, as it does in the reading of linguistic text, where spacing enables a reader to locate sets of letters that constitute words. (Experimental data consistent with this hypothesis are presented in Section 4.) Use of spatial information could facilitate comprehension with either of the kinds of processes that we have considered by focusing attention on the spatial regions that contain characters that are included in constituent units.

We implemented models that use spatial information to restrict generation of cognitive units. The model uses operators and parentheses to form segments of expressions, and only generates term-level units within the segments. Consider the example "5xy+17." Without the process that segments the expression, the model generates units such as 5x, 5y, 5+, xy, x+, and x17. With the segmenter, the units that cross segment boundaries are not generated as terms or constituents of terms. The inclusion of spatial information of this kind in the system made comprehension of expressions considerably more efficient, as would be expected. (By "efficient," in this context we mean that many fewer cycles of activation transfer were used in arriving at a single dominant pattern.)

4. Information-Processing Experiments

We have conducted experiments to investigate information-processing mechanisms involved in comprehending algebra expressions. In one experiment, subjects judged whether a specific operation -- combining terms -- could be applied to expressions. Latencies were measured to test a hypothesis that forming a parsed representation precedes search for individual terms. In two other experiments, subjects were shown algebraically correct and jumbled expressions for brief periods and then were asked about individual characters in the expressions. Their performance provided information about the way in which the structural context of a syntactically correct sequence facilitates comprehension.

Judgments of Applicability*3

The experiment presented a series of expressions. For each expression, the subject's task was to judge whether the operation of combining terms could be applied. The following are examples of the expressions that were used:

(1) 7E - 3M + 9U(4X + 2X)

(2) 7E + 9U(4X + 2P) - 3E

(3) 9U(4E + 2E) + 7E - 3M

(4) 7E - 3M + 9U(4X + 2P)

(5) 7E + 9U(4E + 2P) - 3M

(6) 9E(4E + 2P) + 7E - 3M

The correct response for expressions (1), (2), and (3) is "yes," and the correct response for (4), (5), and (6) is "no."

To respond correctly, the subject must determine both whether there are two or more terms with the same variable and whether the structure of the expression permits their combination. The experiment tested a strong hypothesis about the decision process, namely, that the expression is parsed initially, and a search for combinable terms is restricted to terms that are structurally appropriate -- that is, to single terms at the same level in the expression. This hypothesis was tested with latencies from the negative items. If the process involved a directed search in a parsed expression, these times would all be the same.

The alternative is that the process could be slowed by the presence of like terms in structurally inappropriate locations. This could occur

---

*3. Strauch (1985) provides a more complete report of this experiment.

because of an automatic process of activation, as has been inferred by the existence of a "fan effect" in recognition judgments for sentences (Anderson, 1976). This interpretation would be consistent with the general view expressed in the connectionist models we considered in Section 3. Another interpretation would be use of a strategy by subjects involving discrete processes of first finding like terms, and then determining whether they were in structurally appropriate locations.

16 subjects were recruited from an honors section of calculus at the University of Pittsburgh, to provide a high level of the recognition skills involved. Five blocks of trials were given, with 120 trials in an initial practice block and 120 trials in each of blocks 2-5.

The main finding was that negative expressions with like terms required more time than those without like terms. Considering only blocks 2, 3, 4, and 5, the mean latency for expressions without like terms (e.g., expression (4) above) was 2180 ms, for expressions with two like terms (e.g., (5)) was 2226 ms, and for expressions with three like terms (e.g., (6)) was 2261 ms. The difference between the conditions with like terms and the condition without like terms was significant (95% C.I. = 64 ± 40 ms). The difference between the two conditions with two and with three like terms was not significant (95% C.I. = 35 ± 51 ms), although its direction suggests a graded effect.

The data clearly refute the hypothesis that a parsed representation is formed and searched for combinable terms with consideration only of terms in structurally appropriate locations. Thus, the result is consistent with a process of automatic detection of like terms in a connectionist system, although the data do not rule out a strategic process--either involving systematic search for like terms initially, or using informational results

of an initial connectionistic matching mechanism.

Structural Context in Character Recognition*4

A second question that we have addressed in experiments is the way in which structural context facilitates comprehension of algebra expressions. We have used an experimental method that has been used extensively in studies of word recognition, introduced by Reicher (1969). On each trial the subject is asked about a single character that was presented briefly as part of a larger display. Sometimes the display is a word, and the subject is asked to identify the letter that appeared at one of the positions; two alternative letters are given and the false alternative would also make a word if included with the other letters. On the other trials the letters in the display do not form a word. A robust finding, called the word-superiority effect, is that subjects are better at saying which letter appeared when the context was a word than when it was a nonword. A plausible interpretation (McClelland & Rumelhart, 1981) is that the context of a known word contributes to an activation process that facilitates recognition of the word's individual letters.

The context provided by an expression of algebra depends on structural properties, rather than providing specific known patterns. Our first question, then, was whether recognition of individual characters would be facilitated by the context of a well-formed algebraic expression, compared to a string of characters that is not syntactically correct.

---

*4. Ranney (1985) provides a more complete report of these experiments.

We conducted two experiments. In the first experiment we simply compared algebraic and nonalgebraic contexts and asked subjects which of two letters or which of two numerals appeared at a designated position in the string.

The second experiment asked that question again, and asked another question as well. We also asked whether the context of a well-formed algebra expression would facilitate decisions about the category of a character. The variables in the experiment are shown in Figure 6. On each trial either an algebraic or a nonalgebraic string of characters was displayed, as in the first experiment. There were three kinds of probes, called Same, Different, and Categorical. On Same probes either two letters or two numerals were presented, including the character that appeared at the probed position. On Different probes the correct character was presented along with a character from the opposite category. On Categorical probes, no characters were presented, and the subject just answered whether a letter or a numeral had appeared at the probed position.

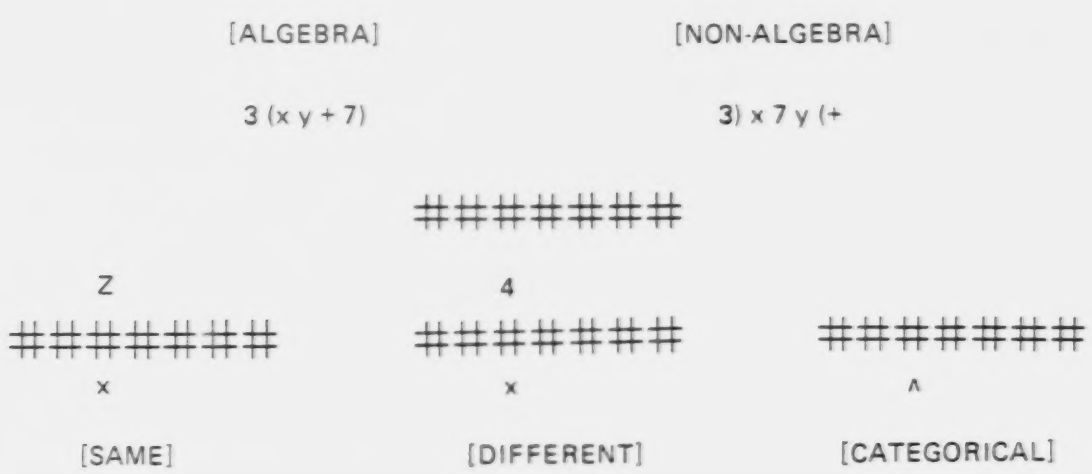

Figure 6. Displays on different types of trials. The sequence was: character string (algebra or non- algebra), then mask, then probe (same-category alternatives, or different-category alternatives, or position only for categorical judgment).

Subjects in the first experiment were 16 volunteers from introductory psychology who were enrolled at the time of the experiment in at least one college mathematics course. The 14 subjects in the second experiment were recruited from an honors calculus section. In the first experiment, data were obtained in two blocks of 84 and 112 trials, preceded by 56 and seven practice trials, respectively. In the second experiment, data were from three 84-trial blocks, preceded by 42, seven, and seven practice trials. During the experiment, exposure durations of the displays were adjusted to maintain a level of approximately 75% responses for each individual subject. Exposure durations were typically in the neighborhood of 100 ms.

The main findings are in Table 2. First, we obtained no facilitation of the recognition of individual characters when the alternatives were in the same category. At least with these materials, structural context did not produce an effect analogous to the word-superiority effect. (For a comparison, we ran a word-superiority experiment with the subjects of the second experiment, using seven-letter words and nonwords and the same displays as were used for algebraic and nonalgebraic strings. A strong word-superiority effect was obtained: .859 correct for words and .679 correct for nonwords. Exposure durations were much shorter than in the algebraic case; no subject's final duration was more than 40 ms.)

Judgments involving categories, however, were facilitated by the structural context of algebraic syntax. Statistically, the main effect of algebra vs. nonalgebra context was significant ($F(1,13) = 43.44$, $p < .0001$), and the two probes involving categories had a significantly

greater effect of display type than the Same probes ($F(1,24) = 5.81$, $p < .025$).

Table 2
Proportions of Correct Response

Probe	Algebra	Non-algebra
Same (Exp. I)	.740	.737
Same (Exp. II)	.752	.741
Different	.815	.760
Categorical	.776	.701

A simple mathematical model was formulated to represent these findings. Assume that when a probe is presented, there is probability R that the subject can recall the character that is probed, and R is independent of the probe condition. If the character is not recalled, but the probe presents alternative characters, there is probability F that the subject can recognize the correct character on the basis of some distinctive feature, presumably orthographic. If the character is not recalled or recognized with a distinguishing feature, or if it is not recalled and only a categorical judgment is requested, then assume that if the display was algebraic there is probability C that the context provides a basis for determining whether the character was a letter or a numeral.

This model implies predictions of proportions correct as follows:

Same/Alg: $R + (1-R)F + .5(1-R)(1-F)$

Diff/Alg: $R + (1-R)F + (1-R)(1-F)C + .5(1-R)(1-C)(1-F)$

Categ/Alg: $R + (1-R)C + .5(1-R)(1-C)$

Same/Nonalg: $R + (1-R)F + .5(1-R)(1-F)$

Diff/Nonalg: $R + (1-R)F + .5(1-R)(1-F)$

Categ/Nonalg: $R + .5(1-R)$

The interesting assumptions are that a single parameter C describes the effect of context on category judgments in both the Different and Categorical probes, and a single parameter F describes the value of distinguishing orthographic features, whether the display was algebraic or nonalgebraic. The model fit the data very well ($X^2(3) = 0.55$, $p > .90$). Maximum-likelihood estimates of the parameters were $R = .40$, $F = .17$, and $C = .25$.

The lack of an effect on recognition of specific characters argues against a model with schematic lexical items like those included in Figure 4. Such a model cannot be ruled out by the data, however, because to have an effect in this experiment would require top-down activation to individual character recognition within the exposure time of about 100 ms, and the effects might be slower than that. Even so, the experiment permitted evidence to support such a model, and that was not obtained.

The facilitating effect of algebraic context on categorical judgments could be caused by general spatial features of the kind that we included in our models that identify segments of characters corresponding to terms. If global spatial features or the locations of operators and parentheses were used to locate segments, then a probe's location would correspond to a position within a segment. Segments often begin with numerals and any

position other than the first must be a letter. Use of this correlation would provide contextual facilitation of the kind that was obtained.

The results obtained with the mathematical model are consistent with a simple hypothesis that the two sources of information involving context and orthographic features are independent. This suggests a model in which general spatial features and features of individual characters are being processed in parallel and without significant interaction.

5. Conclusions

We remark on two implications of our findings for acquisition of the skill of algebra.

First, we believe that the fragmentary character of knowledge of early learners has very serious implications. It does not seem to be a universal characteristic of early stages of skill acquisition -- for example, it does not seem to characterize early knowledge in geometry.

An important problem for theory and for training is to identify characteristics that determine whether early knowledge will be integrated or fragmentary. One possibility is that fragmentary knowledge is likely if the learners are not aware of constraints and goals in the skill domain. If, as seems likely, fragmentary knowledge is not optimal, then attention should be given in designing training to include components that can provide learners with knowledge of the general features of the skill to be acquired. This is consistent with an analysis by Fitts (1962) who noted that successful athletic coaches begin training with a cognitive phase in which they communicate global features of the activities they want their athletes to perform. In related research on algebra, we are exploring

tasks that are designed to provide students with better knowledge of the goals and constraints of algebra procedures.

Secondly, we note that an important component of skill acquisition is learning the information structure of the domain. Our theoretical analysis supports a conjecture that at least some of the errors that are prevalent in early learning indicate a weakness in the learners' ability to represent the materials of the domain -- in the case of algebra, to include structural features in representations of expressions and to include tests for structural features in problem-solving operations that are learned. This is consistent with recent findings in several domains, where it has been found that a lack of ability to represent problem situations adequately is a major source of difficulty in problem solving of novices (e.g., Chi, Feltovich & Glaser, 1981; Heller & Reif, 1984; Riley, 1984; Riley, Greeno & Heller, 1983). More attention to the skills and knowledge needed to represent problems, and training materials specifically focused on representational ability, may be an important general suggestion that emerges from recent cognitive studies of training, including those described in this report.

REFERENCES

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review, 84, 413-451.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Anderson, J. R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.

Brown, J. S. & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

Brown, J. S. & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. Cognitive Science, 4, 379-426.

Carry, L. R., Lewis, C., & Bernard, J. E. (1980). Psychology of equation solving: An information processing study (Tech. Rep.). Austin: University of Texas, Department of Curriculum & Instruction.

Chaiklin, S., & Lesgold, S. B. (1984). Prealgebra students' knowledge of algebraic tasks with arithmetic expressions (Tech. Rep. No. UPITT/LRDC/ONR/APS-16). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge MA: MIT Press.

Davis, R. B., Jockusch, E., & McKnight, C. (1978). Cognitive processes in learning algebra. Journal of Children's Mathematical Behavior, 2, 5-320.

Dixon, T. R., & Horton, D. L., (Eds.). (1968). Verbal behavior and general behavior theory. Englewood Cliffs, NJ: Prentice-Hall.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. Cognitive Science, 6, 205-254.

Fitts, P. M. (1962). Factors in complex skill training. In R. Glaser (Ed.), Training research and education (pp. 177-197). Pittsburgh: University of Pittsburgh Press.

Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 1, pp. 13-75). Hillsdale, NJ: Erlbaum.

Groen, G. J., & Resnick, L. B. (1977). Can preschool children invent addition algorithm? Journal of Educational Psychology, 69, 645-652.

Heller, J. I. & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. Cognition and Instruction, 1, 117-175.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), Parallel models of associative memory (pp. 161-187). Hillsdale, NJ: Erlbaum.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems (pp. 25-50). New York: Academic Press.

McClelland, J. S., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. Communications of the ACM, 19, 113-126.

Ranney, M. A. (1985). The role of structural context in perception: Syntax in the recognition of algebraic expressions. Manuscript submitted for publication.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81, 274-280.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.), The development of mathematical thinking (pp. 153-196). New York: Academic Press.

Riley, M. S. (1984). Structural understanding in performance and learning. Unpublished doctoral dissertation, University of Pittsburgh.

Sacerdoti, E. D. (1977). A structure for plans and behavior. New York: Elsevier-North Holland Publishing Co.

Sleeman, D. H., & Smith, M. J. (1981). Modelling students' problem solving. Artificial Intelligence, 16, 171-187.

Strauch, C. (1985). The effect of irrelevant terms on judgments of combinability. Unpublished manuscript, University of Pittsburgh, Learning Research and Development Center.

Van Lehn, K. (1983). Felicity conditions for human skill acquisition: Validating an AI-based theory (Tech. Rep. No. CIS21). Palo Alto, CA: Xerox PARC.

Van Lehn, K. (1981). Bugs are not enough: Empirical studies of bugs, impasses, and repairs in procedural skills (Tech. Rep. No. CIS-11). Palo Alto, CA: Xerox PARC.